

Trabajo de Fin de Grado

Integración de sistemas para el análisis de redes sociales utilizando procesamiento de lenguaje natural.

Constanza Polette León Baritussio

La Laguna, 18 de Abril de 2017

D. Dagoberto Castellanos Nieves, con N.I.F. 79234766-L profesor Ayudante Doctor de Universidad adscrito al Departamento de Ingeniería Informática y de Sistemas de la Universidad de La Laguna, como tutor.

C E R T I F I C A (N)

Que la presente memoria titulada:

“Integración de sistemas para el análisis de redes sociales utilizando procesamiento de lenguaje natural.”

ha sido realizada bajo su dirección por D. Constanza Polette León Baritussio, con N.I.F. X-6337312-F.

Y para que así conste, en cumplimiento de la legislación vigente y a los efectos oportunos ﬁrman la presente en La Laguna a 18 de Abril de 2017.

Agradecimientos

XXXcgghhh

XXX

XXX

XXX

Licencia

\* Si NO quiere permitir que se compartan las adaptaciones de tu obra y NO quieres permitir usos comerciales de tu obra indica:



© Esta obra está bajo una licencia de Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional.

\* Si quiere permitir que se compartan las adaptaciones de tu obra mientras se comparta de la misma manera y NO quieres permitir usos comerciales de tu obra indica:

C:\Users\Usuario\Desktop\memoriaTFG_1415_LaTeX\tfg\images\by-nc-sa_88x31.png

© Esta obra está bajo una licencia de Creative Commons Reconocimiento-NoComercial-CompartirIgual 4.0 Internacional.

\* Si quiere permitir que se compartan las adaptaciones de tu obra y NO quieres permitir usos comerciales de tu obra indica:

C:\Users\Usuario\Desktop\memoriaTFG_1415_LaTeX\tfg\images\by-nc_88x31.png

© Esta obra está bajo una licencia de Creative Commons Reconocimiento-NoComercial 4.0 Internacional.

\* Si NO quiere permitir que se compartan las adaptaciones de tu obra y quieres permitir usos comerciales de tu obra indica:

C:\Users\Usuario\Desktop\memoriaTFG_1415_LaTeX\tfg\images\by-nd_88x31.png

© Esta obra está bajo una licencia de Creative Commons Reconocimiento-SinObraDerivada 4.0 Internacional.

\* Si quiere permitir que se compartan las adaptaciones de tu obra mientras se comparta de la misma manera y quieres permitir usos comerciales de tu obra (licencia de Cultura Libre) indica:

C:\Users\Usuario\Desktop\memoriaTFG_1415_LaTeX\tfg\images\by-sa_88x31.png

© Esta obra está bajo una licencia de Creative Commons Reconocimiento-CompartirIgual 4.0 Internacional.

\* Si quiere permitir que se compartan las adaptaciones de tu obra y quieres permitir usos comerciales de tu obra (licencia de Cultura Libre) indica:

C:\Users\Usuario\Desktop\memoriaTFG_1415_LaTeX\tfg\images\by_88x31.png

© Esta obra está bajo una licencia de Creative Commons Reconocimiento 4.0 Internacional.

**Resumen**

El objetivo de este trabajo ha sido, desarrollar una integración de sistemas, en concreto un sistema de alerta y un sistema de análisis de sentimiento.

La competencia [E6], que figura en la guía docente, indica que en la memoria del trabajo se ha de incluir: antecedentes, problemática o estado del arte, objetivos, fases y desarrollo del proyecto, conclusiones, y líneas futuras.

El documento de memoria debe tener un máximo de 50 páginas.

No se deben dejar páginas en blanco al comenzar un capítulo, ya que el documento no está pensado para sea impreso sino visionado con un lector de PDFs.

También es recomendable márgenes pequeños ya que, al firmar digitalmente por la Sede, se coloca un marco alrededor del texto original.

El tipo de letra base ha de ser de 14ptos.

**Palabras clave:** Red Social, similitud semántica.

**Abstract**

*The objective of this work has been,the extraction of data from the social network Twitter,it is a perfoming tool capable of analytics on the semantic similaritis between the concepts used on social networks.*

***Keywords:*** *Social network, semantic similarity.*

Índice General

[Capitulo 1. Introducción 10](#_Toc479411507)

[1.1 Internet 10](#_Toc479411508)

[1.2 Web 2.0 10](#_Toc479411509)

[1.3 Redes Sociales 11](#_Toc479411510)

[1.4 Redes Colaborativas 16](#_Toc479411511)

[1.4.1 Introducción y definición 16](#_Toc479411512)

[1.4.2 Tipologías de redes organizacionales 17](#_Toc479411513)

[1.4.3 Similitudes con las redes sociales 18](#_Toc479411514)

[1.5 Tecnologías 19](#_Toc479411515)

[1.5.1 Introducción 19](#_Toc479411516)

[1.5.2 Procesamiento del lenguaje natural (PLN) 19](#_Toc479411517)

[1.5.3 RStudio. 22](#_Toc479411518)

[1.5.4 Lenguaje R 22](#_Toc479411519)

[1.6 Red social del proyecto. 24](#_Toc479411520)

[Capítulo 2. Sistemas de alerta y sistemas de análisis de sentimientos. 25](#_Toc479411521)

[Capítulo 3. 26](#_Toc479411522)

[Integración de los sistemas de alerta y sistemas de análisis de sentimientos propuestos. 26](#_Toc479411523)

[1.7 Primer apartado de este capítulo 26](#_Toc479411524)

[1.8 Segundo apartado de este capítulo 26](#_Toc479411525)

[1.9 Tercer apartado de este capítulo 26](#_Toc479411526)

[Capítulo 4. Recomendaciones o conclusiones. 27](#_Toc479411527)

[Conclusiones y líneas futuras 28](#_Toc479411528)

[Summary and Conclusions 29](#_Toc479411529)

[1.10 First Section 29](#_Toc479411530)

[Presupuesto 30](#_Toc479411531)

[1.11 Sección Uno 30](#_Toc479411532)

[Apéndice A. Título del Apéndice 1 31](#_Toc479411533)

[A.1. Algoritmo XXX 31](#_Toc479411534)

[A.2. Algoritmo YYY 31](#_Toc479411535)

[Apéndice B. Título del Apéndice 2 33](#_Toc479411536)

[B.1. Otro apéndice: Sección 1 33](#_Toc479411537)

[B.2. Otro apéndice: Sección 2 33](#_Toc479411538)

[Bibliografía 34](#_Toc479411539)

Índice de figuras

[Figura 1. Comparativa Deportes 15](#_Toc479412315)

[Figura 2. Comparativa Universidades 15](#_Toc479412316)

[Figura 3.Comparativa Gobierno 16](#_Toc479412317)

Índice de tablas

[Tabla 1.1. Esta es la primera tabla con estilo **Tabla 1** (usar estilo **Celda de tabla** y luego dar formato). El encabezado tiene estilo **Encabezado de tabla**. 3](#_Toc418604518)

[Tabla 1.2. Segunda tabla con estilo **Tabla 2** (ídem). 3](#_Toc418604519)

[Tabla 7.1. Tabla resumen de los Tipos. 3](#_Toc418604520)

## Capitulo 1. Introducción

La memoria está compuesta por los siguientes cuatro capítulos: 1. Introducción, en este capítulo se describen los conceptos clave, como: Internet, Web 2.0, redes sociales, redes colaborativas. Tecnologías necesarias para llevar a cabo el proyecto: Procesamiento del lenguaje natural, entornos de desarrollo como Rstudio, lenguaje de programación R y sus librerías. 2. Sistemas de alerta y sistemas de análisis de sentimientos, pretende definir en qué consiste cada uno de estos y exponer ejemplos sobre en que ámbitos se están desarrollando y su utilidad. 3. Integración de sistemas de alerta y sistemas de análisis de sentimientos, la intención de este capítulo es explicar cómo y con qué metodologías, se realizo la elaboración de una herramienta, que permite la integración de un sistema de alerta y un sistema de análisis de sentimientos. 4. Recomendaciones y conclusiones, con este capítulo se recogen las posibles mejoras que se podrían implementar para tener una herramienta más optima, así como las conclusiones obtenidas después de la realización del proyecto.

### Internet

El nombre Internet proviene de las palabras en inglés Interconnected Networks, que significa "redes interconectadas". Sus orígenes se remontan a 1969, cuando se estableció la primera conexión entre ordenadores, conocida como ARPANET.

Internet, es una arquitectura de sistema que ha revolucionado las comunicaciones y los métodos de comercio al permitir que varias redes de computadoras estén interconectadas. Proporciona un mecanismo potente y accesible por cualquier individuo, ya que, ofrece una amplía información que puede ser utilizada para cualquier propósito en general. La difusión de textos digitalizados, imágenes y grabaciones de audio y vídeo, ha dado lugar a una explosión de información [1].

El internet, se ha convertido en uno de los medios más importantes desde finales del siglo 20 y principios del 21 en la sociedad, ya que, trajo consigo una nueva forma de comunicación entre las personas y una mejora en la transmisión y acceso de la información eliminando así las barreras de espacio/tiempo, surgiendo así la World WideWeb (Tim Berners-Lee y Robert Cailliau 1990) o también conocida como Web 1.0. Mediante el concepto denominado la “webvolución” (Kapp y O’Driscoll, 2010), en otras palabras, la evolución de la web, el enfoque del concepto de la web ha ido transformándose de “acceso y hallazgos” (Web 1.0), a “compartir, participar y colaborar”(Web 2.0), hasta llegar a la “colaboración inmersiva y co-creación”(Web 3.0) (Domínguez y Llorente, 2009);Llorente, 2013ª).

### Web 2.0

Después de la llegada de la World Wide Web en los 90 y la Web2.0 (Tim O´Reilly y Dale Dougherty en 2004), para referirse a una segunda generación en la historia de la Web. La Web 2.0 se puede entender cómo –“todas aquellas utilidades y servicios de Internet que se sustentan en una base de datos, la cual puede ser modificada por los usuarios del servicios, ya sea en su contenido (añadiendo, modificando o borrando información o asociando datos a la información existente), pues bien en la forma de presentarlos, o en contenido y forma simultáneamente”- (Ribes, 2007). Se puede afirmar que el término “Web 2.0” representa la evolución de las aplicaciones tradicionales hacia aplicaciones web enfocadas en el usuario final. No se trata de una nueva tecnología sino de una actitud de colaboración y participación de las personas para proporcionar mejores datos, nuevos servicios y aplicaciones on-line [2]. En donde los usuarios pueden colaborar e interactuar entre sí [3], creando contenidos dentro de una comunidad virtual y conectándose con otros individuos.

Aunque todos los sitios de la Web están interconectados entre sí, con el fin de organizar los recursos digitales (Cobo y Pardo 2007)[4],proponen una taxonomía compuesta por cuatro líneas fundamentales que ayudará al usuario a seleccionar aquellas aplicaciones que cubran de una mejor manera sus necesidades e intereses. La estructura propuesta se denomina “los cuatro pilares de la Web 2.0”, que son:

* **Redes sociales:** describe todas aquellas herramientas diseñadas para la creación de espacios que promuevan o faciliten la formación de comunidades e instancias de intercambio social.
* **Contenidos:** hace referencia a aquellas herramientas que favorecen la escritura en línea, así como su distribución e intercambio.
* **Organización social e inteligente de la información:** herramientas y recursos para etiquetar, sindicar e indexar, que faciliten el orden y almacenamiento de la información, así como otros recursos disponibles en la Red.
* **Aplicaciones y servicios (mashups):** dentro de esta clasificación se incluye numerosas herramientas, software, plataformas en línea y un hibrido de recursos creados para ofrecer servicios de valor al usuario final.

### Redes Sociales

Como se describía en la sección anterior, la Web 2.0 está compuesta de comunidades de usuarios y una gama especial de servicios, como por ejemplo, las redes sociales como Linkedin, Instagram, los blogs como Blogger, Wordpress y las wikis como Wikipedia.

Para definir a las redes sociales, existen diferentes prospectivas; las que se centran en la comunicación y colaboración o las que se basan en estudio de las características estructurales de sus herramientas (Cabero,Barroso, Llorente y Marín,2013). Algunos autores como Boyd y Ellison(2008), definen a las redes sociales como herramientas que facilitan el construir un perfil público dentro de un sistema delimitado, que les permite articular una lista de diferentes usuarios con los que comparte una conexión, y además poder ver las conexiones que suceden dentro de su sistema. Otros autores como Haro (2010), las define como una estructura social que se puede representar como nodos conectados por aristas, en donde los nodos representan a los individuos y las aristas las relaciones existentes entre ellos. Para el autor Cabero (2012), se centra en un entorno educativo, señala que son un entorno que permite la discusión de diferentes problemáticas en espacios interactivos y flexibles de aprendizaje. Nosotros definiremos a las redes sociales como una estructura compuesta por un grupo de individuos u organizaciones y que están relacionados entre sí, ya sea, mediante relaciones laborales, amistad, familiar. Además de un punto de encuentro en donde los usuarios comparten, crean y acceden a la información.

En la sociedad actual, las redes sociales despiertan un inusitado interés social y profesional. Así como en el ámbito académico el interés se centra en conocer los efectos que estas provocan en el éxito y desarrollo educativo de cualquier estudiante-más si cabe, niveles universitarios-(Abramsn, 2011) [5]. En el resto de ámbitos como el laboral, social y relacionamiento, la creación de las redes sociales supone una nueva forma de comunicación. Estas herramientas ofrecen un espacio virtual para que, miles de personas pueden compartir contenidos multimedia con intereses similares (Granovetter, 1973) [6]. Se consideran puntos de encuentro, en los que es posible acceder a la información, compartir impresiones, consultar archivos y recursos disponibles a tiempo real (Nass De Ledo, 2011) [7].

Redes sociales como Facebook, Twitter, Orkut, son estructuras compuestas por una infinidad de usuarios relacionados entre sí, independientemente que se conozcan o no. En estas redes sociales, al tener un gran número de usuarios, se manejan grandes volúmenes de datos, ya que, los individuos comparten, buscan y difunden información a través de texto, imágenes o audio. Gracias a las redes sociales, las personas pueden expresar opiniones digitales en foros, debates y hacer que estas lleguen a un público muy amplio y diverso.

En el año 2002, comenzaron a aparecer los primeros sitios web, promocionando las redes de círculos de amigos en línea. La popularidad de estos sitios creció rápidamente y compañías como por ejemplo, Myspace y Linkedin surgieron en el año 2003, Google lanzó Orkut en 2004 y en el mismo año se dio a conocer Facebook, en 2006 se lanzó Twitter.

Uno de los objetivos de la creación de estas redes, es crear un sitio de interacción virtual en donde miles de personas de todo el mundo puedan compartir, comunicarse y cooperar entre sí.

Clasificación de las redes sociales.

Las redes sociales según su tipología pueden clasificarse entre redes sociales horizontales o generales y verticales o especializadas.

Las redes sociales horizontales, no tienen una temática definida, están orientadas al público en general y se centran en las relaciones de contacto entre usuarios. Provee herramientas para que las personas se relacionen entre sí y todas comparten características: crear un perfil, compartir contenidos, generar lista de contactos [8]. Algunas de ellas son:

* Facebook.
* Orkut.
* Google+.

En cuanto a las redes sociales verticales, tienden a especializarse según los intereses de sus usuarios, está orientada a un público específico y diariamente ganan miles de seguidores. Cualquier red social que se centre en un tema en concreto, puede dar origen a una nueva categoría de clasificación y es por ello que destacaremos las más relevantes, como por ejemplo las redes sociales según su temática, actividad y por ultimo según el contenido compartido [9].

1. **Temática:**
   1. Profesionales: Son redes enfocadas en negocios y actividades comerciales. Los usuarios definen un perfil en donde detallan su ocupación, estudios académicos, experiencia laboral. La más conocida es Linkedin.
   2. Viajes: Estas redes sociales, conectan viajeros que comparten experiencias, sirviendo así de guía para otros viajeros. Un ejemplo de ellas sería Travellerspoint, tripadvisor.
   3. Relacionamiento: Son redes sociales propias para relaciones personales de pareja, por ejemplo: meetic.
2. **Actividad:**

* Microblogging: Estas redes sociales ofrecen un servicio de envío y publicación de mensajes breves de texto. También permiten seguir a otros usuarios, aunque esto no establece necesariamente una relación recíproca, como los seguidores. La red social más destacada es Twitter.
* Geolocalización: Permiten mostrar la localización de una persona o establecimiento. Mediante estas redes sociales los usuarios pueden localizar el contenido digital que comparten. Por ejemplo: Foursquare, Trippy.

1. **Contenido:**
   1. Fotos: Estos servicios ofrecen la posibilidad de almacenar, ordenar, buscar y compartir fotografías. Las más importantes en número de usuarios son: Instagram, Flickr, Fotolog.
   2. Vídeos: Los sitios web de almacenamiento de vídeos se han popularizado de tal manera que en los últimos años incorporan la creación de perfiles y listas de amigos para la participación colectiva mediante los recursos de los usuarios, y los gustos sobre los mismos. Algunos son: Youtube, Dailymotion.
   3. Documentos: Por la red navegan documentos de todo tipo en formatos diversos, en estas redes sociales podemos encontrar, publicar y compartir los textos definidos por nuestras preferencias de una manera fácil y accesible. Su mayor exponente es Scribd, GoogleDrive.

Nosotros haremos una clasificación gráfica para representar a las redes sociales más importantes en la red. Poniendo énfasis en lo que respecta a la compartición de información mediante la opinión de los usuarios. Y a los temas tratados con especial relevancia como por ejemplo temas de actualidad como la política, Internet y telecomunicaciones, destinos turísticos, noticias en general, deportes y así una infinidad de temáticas.

Según las estadísticas del último año en todo el mundo a lo que deportes se refiere, Twitter es la más utilizada a la hora de divulgar opiniones acerca de los partidos, competiciones y campeonatos.



Deportes

Figura 1. Comparativa Deportes

En el caso de las facultades y universidades que utilizan redes sociales, la más destacada es Twitter frente a redes sociales como Orkut o Instagram.



Facultades

Y

Universidades

Figura 2. Comparativa Universidades

Comparativa en España de las redes de mensajería rápida más utilizadas, a la hora de compartir contenidos a cerca del gobierno español.

[](https://trends.google.es/trends/explore?cat=19&date=today%203-m&geo=ES&q=twitter,telegram,/m/0gwzvs1)

Gobierno

Figura 3.Comparativa Gobierno

### Redes Colaborativas

#### Introducción y definición

Anteriormente se ha definido el concepto de redes sociales y sus clasificaciones. En este apartado se definirá que es una red colaborativa y su similitud con respecto a las redes sociales.

Entre los varios tipos de redes que existen, las Redes Colaborativas (RC) tienen una especial relevancia. Se puede definir a una Red Colaborativa como una red compuesta por varias entidades autónomas, dispersas geográficamente y distintas en términos de: entorno, cultura, capital social y objetivos. Colaboran para alcanzar metas comunes y su interacción se soporta en redes de ordenadores (Camarinha-Matos, 2005). Según (Camarinha-Matos, 2005), en las RC, la colaboración entre los miembros de la red hace posible alcanzar los objetivos marcados y que de otra manera independiente estos objetivos serían más costos de alcanzar para las organizaciones. Otros autores como (Brass et al., 2004), describe a las RC como, una red que se compone de un conjunto de vínculos que pueden representar relaciones, o ausencia de ellas.los nodos pueden ser actores, unidades organizativas y organizaciones. El objeto de intercambio define el tipo de red a estudiar (bienes, servicios, influencias, afecto).Las relaciones se asumen como perdurables [10].

Podemos definir a una Red Colaborativa como, una estructura compuesta por personas u organizaciones que colaboran entre sí para lograr objetivos comunes, independientemente del ámbito en donde se desarrollen, en donde interactúan mediante los ordenadores. La colaboración entre los componentes de las organizaciones, es fundamental para el desarrollo de las redes organizacionales. En las RC, el proceso colaborativo se compone de una colaboración, cooperación, coordinación y networking, que son realizadas mediante actividades en conjunto para así lograr objetivos comunes.

#### Tipologías de redes organizacionales

En las redes colaborativas, según (Aldrich y Wheteh 1981), identifican cuatro escenarios posibles en los que analizar los patrones de relaciones entre organizaciones, clasificados en:

1. **Análisis centrado en una organización:**
   1. **Relación binaria**: el análisis de la relación se realiza respecto a dos entidades que establecen una acuerdo de colaboración. El contenido de la relación se centra en definir:
      1. Predisposición, o no, para actuar de manera conjunta, mediante el intercambio de bienes o servicios.
      2. Dependencia mutua. Resultado de la acción conjunta de las dos organizaciones.
      3. Organización del vínculo. Representa el grado de colaboración que puede medirse en términos de: formalización del acuerdo, intensidad, reciprocidad y estandarización.
   2. **Relación estelar**: Es un modelo centralizado en el que la organización que predomina, analiza de manera dual, su actividad de cooperación y colaboración con cada uno de las otras organizaciones o entidades.
2. **Análisis Orientado a la red:**
   1. **Red de trabajo:** se trata de un conjunto de organizaciones que se movilizan de manera cooperativa con la finalidad de alcanzar un objetivo determinado aunque para lograrlo no necesariamente deben adquirir una entidad única (Borgatti, 2003). La duración de esta colaboración, depende del objetivo consensuado y predeterminado cuyo alcance de manera individual no sería posible. Este tipo de colaboración, no se enfoca hacia las relaciones entre los participantes, sino en el comportamiento organizativo global. Para conseguirlo, los participantes deben determinar:
      1. El número de actores implicados.
      2. El ámbito de poder al que accede la organización escogida como líder
      3. La similitud de valores y actitudes entre los miembros
      4. El efecto que tiene el comportamiento de otros equipos frente a éste.
   2. **Entre Redes:** la colaboración entre redes, representa a un conjunto de redes, generalmente con un número elevado de integrantes, que mantienen algún tipo de relación entre ellas (Borgatti, 2003). Difiriendo con la red de trabajo, la colaboración entre redes, posee una naturaleza dinámica y su extensión temporal puede ser indefinida [11].

#### Similitudes con las redes sociales

Anteriormente hemos definido a las redes sociales como una estructura compuesta por un grupo de individuos u organizaciones y que están relacionados entre sí, y en donde los usuarios comparten, crean y acceden a la información. Y una red colaborativa como una estructura compuesta por personas u organizaciones que colaboran entre sí ya sea en algún proyecto, investigación para lograr objetivos, comunes, independientemente del ámbito en donde se desarrollen, en donde interactúan mediante los ordenadores.

Al igual que las redes colaborativas, las redes sociales son unas estructura organizada que puede estar compuesta por individuos u organizaciones. En donde podemos representar su tipología mediante nodos conectados entre sí, como ocurre con las redes de trabajo o la relación entre redes. También nos encontramos que en las redes sociales, como es el caso de los Blogs y las Wikis los usuarios trabajan de manera colaborativa porque al compartir la información, permite que otros usuarios accedan a ella creándose así un entorno colaborativo en el que poder ayudarse mutuamente compartiendo conocimientos.

### Tecnologías

#### Introducción

En secciones anteriores, se ha definido a las redes sociales, y también como estas están compuestas por miles de usuarios. Estas redes sociales al tener un gran número de usuarios, están constantemente compartiendo y difundiendo información a través de imágenes, texto o audio. Por lo que se generan grandes volúmenes de datos. Para poder analizar esa gran cantidad de datos, es necesario la utilización de tecnologías como: el procesamiento del lenguaje natural (PLN), entornos de desarrollo como RStudio y lenguajes de programación como el lenguaje R.

#### Procesamiento del lenguaje natural (PLN)

Para entender el procesamiento del lenguaje natural, primero definiremos en que consiste el lenguaje natural (LN). Un lenguaje natural es aquel que ha evolucionado con el tiempo para fines de comunicación humana, como el español o alemán [12]. El lenguaje natural, es la vía de comunicación que utilizan las personas para comunicarse entre ellas habitualmente.

El Procesamiento de Lenguaje Natural (PLN), se entiende como la habilidad que tiene una máquina para procesar la información comunicada, no simplemente las letras o los sonidos del lenguaje.Un traductor autómatico sería un ejemplo de dispositivo o software de PLN [13].

El PLN tiene aplicaciones muy variadas, ya que su alcance es muy extenso, algunas aplicaciones son:

* Traducción automática.
* Recuperación de la información.
* Extracción de Información y Resúmenes.
* Tutores inteligentes.
* Reconocimiento de Voz.

La arquitectura de un sistema de PLN, se fundamenta en una definición del LN por niveles, que son: fonológico, morfológico, sintáctico, semántico y pragmático. En este trabajo, nos centraremos en los niveles: sintácticos, semántico y pragmático.

* **Nivel sintáctico:** trata de cómo las palabras pueden unirse para formas oraciones, fijando el papel estructural que cada palabra juega en la oración y que sintagmas son parte de otros sintagmas.
* **Nivel semántico:** trata del significado de las palabras y de cómo los significados se unen para dar significado a una oración, también se refiere al significado independiente del contexto, es decir de las oraciones aisladas.
* **Nivel Pragmático:** trata de cómo las oraciones se usan en distintas situaciones y de cómo el uso afecta al significado de las oraciones. Se reconoce un subnivel recursivo: discursivo, que trata de cómo el significado de una oración se ve afectado por las oraciones inmediatamente anteriores [14].

Según el autor (Alexander Gelbukh, 2010), existen cuatro componentes necesarios para usar y manejar de manera más provechosa el conocimiento plasmado en textos. Estos son: la digitalización, la recuperación, la presentación de la información y su uso directo por el software.

1. **Digitalización de la información:** Existe muchísimos documentos, libros, manuales en donde su contenido está plasmado en papel y es difícil de acceder a la información. Por ello es importante que a la hora de usar el PLN el contenido este en formato digital. Por digitalización podemos entender como la obtención de la información en formato digital en donde el texto es una secuencia de letras, por ejemplo los correos electrónicos, documentos en periódicos digitales, información en las redes, etc.
2. **Recuperación de la información**: también conocida como búsqueda de la información, tiene la problemática que un mismo concepto o idea puede expresarse de maneras distintas, utilizando diferentes palabras. Por ejemplo: el usuario expresa su interés con la frase “la derrota del Madrid” y el documento relevante para tal petición es “la victoria del Barcelona”. Los dos textos no tienen ninguna palabra en común, pero un humano, usando su experiencia lingüística (derrota—victoria) y su conocimiento del mundo (Madrid—Barcelona) fácilmente detectaría la relevancia del documento para la petición. Se han alcanzado avances significativos para que los programas puedan usar este tipo de razonamiento para así, satisface de una mejor manera las necesidades de los usuarios.
3. **Presentación de la información**: es importante que la información contenida en los textos, este representados de una manera eficaz. Un ejemplo sería, la construcción automática de resúmenes, es decir, un generador automático de resúmenes, a partir de un texto largo, detectar lo más transcendente y presentarlo en un texto corto que se pueda leer en un tiempo más razonable. Otra forma de resumir la información para hacerlos más manejables sería, agrupar y clasificar los documentos que sean semejantes entre sí. La minería de texto es otra manera de resumir el contenido textual (encontrar opiniones expresadas en los textos, las tendencias de cambio de estas opiniones o las relaciones inesperadas entre los eventos descritos en los textos), la extracción de información (llenar bases de datos sobre un tema específico, leyendo los textos) y sistemas de soporte a la toma de decisiones (buscar, sintetizar y presentar de manera eficiente la información relevante para un directivo).
4. **Uso de la información**: el uso de la información contenida en los textos por el mismo software para resolver tareas más complejas, permite a la máquina encontrar el conocimiento necesario de los textos disponibles. Por ejemplo: los artículos científicos o los libros de texto.

#### RStudio.

Rstudio es un entorno de desarrollo integrado (IDE) para R. Incluye una consola, un editor de resaltado de sintaxis, que soporta la ejecución de código directo, así como herramientas para: la depuración y gestión de espacio de trabajo. Rstudio está disponible para: Windows, Mac y Linux se puede ejecutar en el escritorio o en un navegador conectado a rstudio Server o rstudio Server Pro (Debian / Ubuntu, RedHat / CentOS, y SUSE Linux)[15].

RStudio tiene como objetivo proporcionar un entorno de desarrollo informático estadístico R. Permite el análisis y desarrollo para que se lleven a cabo los análisis correspondientes con el lenguaje R. Las principales características son:

* Acceso Rstudio localmente.
* El resaltado de sintaxis, completado de código, y la sangría inteligente.
* Ejecutar código R directamente desde el editor de código fuente.
* Saltar rápidamente a la definición de funciones.
* Gestionar fácilmente múltiples directorios de trabajo utilizando proyectos.
* R ayuda integrada y documentación.
* Depurador interactivo para diagnosticar y corregir los errores rápidamente.
* Paquete de herramientas de desarrollado por usuarios.

#### Lenguaje R

Introducción.

El lenguaje R además de ser un lenguaje de programación, también es un entorno con un enfoque al análisis s y gráficos. Es un proyecto GNU similar al lenguaje de programación S. Fue desarrollado en Bell Laboratories anteriormente AT & T, ahora Lucent Technologies), por John Chambers y colegas. R puede considerarse como una implementación diferente a S. Entre estos dos lenguajes de programación, existen algunas diferencias importantes, aunque mucho código escrito para S se ejecuta bajo R.

R proporciona una gran variedad de modelos estadísticos ( modelos lineales y no lineales, pruebas estadísticas clásicas, clasificaciones, agrupaciones, etc.) y técnicas graficas además es muy extensible. El lenguaje R proporciona facilidad a la hora de crear publicaciones bien diseñadas incluyendo símbolos matemáticos y formulas donde se es necesario. Además es uno de los lenguajes más utilizados por la comunidad de programadores, y es muy popular en campos como: la minería de datos, investigaciones biomédicas, la bioinformática, y finanzas. R está disponible como software libre bajo la licencia de GNU, compila y se ejecuta en una amplia variedad de plataformas como: UNIX, Windows, MacOS. Ofrece la posibilidad de trabajar con distintas bibliotecas o librerías que ofrecen funciones para analizar los datos que se le provee [16].

Características de R.

R es un conjunto integrado de paquetes para la manipulación de datos, el cálculo y la visualización grafica[17].Incluye:

* Una instalación eficaz de tratamiento y almacenamiento de datos.
* Una serie de operadores para cálculos matemáticos.
* Una gran colección de herramientas integradas para el análisis de datos.
* Instalaciones gráficas para el análisis de datos y visualización.
* Proporciona gráficas para análisis de datos, que funcionan directamente sobre pantalla o impresora.
* Un lenguaje de programación bien desarrollado, simple y efectivo, que incluye condicionales, ciclos, funciones recursivas y posibilidad de entradas y salidas.
* Permite que los usuarios lo extiendan definiendo sus propias funciones.
* Orientado a objetos, característica que hereda de S.

Librerías.

R es miembro de un proyecto colaborativo y abierto, en donde sus usuarios pueden publicar paquetes y extender su configuración básica. R se puede extender fácilmente con los paquetes, existen unos ocho paquetes suministrados con la distribución de R y muchos de ellos están disponibles en CRAN (<https://www.r-project.org/>).

Librerías para el proyecto:

En nuestro proyecto para su la elaboración utilizaremos algunas librerías como:

* **library (“twitteR”):** Proporciona una interfaz a la API web de Twitter.
* **library(“tm”):** Proporciona un marco para aplicaciones de minería de texto dentro de R.
* library(“ROAuth”): Proporciona una interfaz de OAuth 1.0, permitiendo asi a los usuarios de Twitter autenticarse vía OAuth.
* library(“streamR”): Este paquete ofrece una serie de funciones que permiten a los usuarios acceder por filtro a los usuarios de Twiter.

### Red social del proyecto.

En nuestro proyecto tras analizar las distintas redes sociales, nos hemos decantado por trabajar con la red social Twitter. Al tener miles de usuarios compartiendo información diariamente a través de los tweets, hashtags, entre otros, se manejan grandes volúmenes de datos con diversidad de ámbitos como: la política, el turismo, opiniones digitales, etcétera. Para abarcar el procesamiento de todos los datos generados en la redo social Twitter utilizaremos el lenguaje R, ya que, es un lenguaje muy potente a la hora de tratar con la minería de datos. También trabajaremos con procesamiento del lenguaje natural (PLN), para poder analizar los datos extraídos de la red social. Utilizando mecanismos como los sistemas de alerta y sistemas de análisis de sentimientos para el tratamiento de los datos.

## Capítulo 2. Sistemas de alerta y sistemas de análisis de sentimientos.

### Sistemas de alerta.

Un sistema de alerta, se podría definir como: sistema que ante uno o varios eventos, es capaz de generar una notificación o alerta, para indicar a un usuario o grupo de usuarios sobre un suceso, acontecimiento que está ocurriendo en ese momento o simplemente retornar una información.

Los sistemas de alerta han sido implementados en diversos campos como: la vigilancia tecnológica, la medicina, la climatología, la alimentación, entre otros.

#### Sistemas de alertas de vigilancia tecnológica

En un sistema de alertas de vigilancia tecnológica, como por ejemplo “ALERTAS”, ofrece información de actualidad sobre: ciencia, tecnología e innovación. A través de estas alertas, los usuarios pueden tener acceso a: eventos actualizados, noticias diarias, artículos recientes, ferias y congresos. Este sistema para poder ofrecer el servicio de alerta, necesita que el usuario este registrado y que además en el formulario de registro, especifique sus preferencias. Posteriormente el sistema, recogerá está información para llevar a cabo y de manera diaria el proceso de detección, clasificación y priorización de resultados para así, ofrecer al usuario un resumen de noticias, eventos y novedades.

El sistema ALERTAS incorpora tecnologías del lenguaje humano para la recuperación, extracción y clasificación de información documental, desarrolladas bajo los estándares de la web semántica. Con la suma de estrategias y técnicas es posible presentar al usuario solo aquella información que cubra sus necesidades. Estas son: la creación y mantenimiento de ontologías, la asignación semi-automática de metadatos y generación de vinculaciones y relaciones semánticas entre estos para la extracción de información textual .

ALERTAS, permite la clasificación de la información procedente de fuentes relevantes en el ámbito de la científico. Extrae solamente la información relevante dependiendo de las preferencias de los usuarios y la almacena en bases de datos para su posterior procesamiento. Esta compuesto por varias partes [18]:

* Clasificación inteligente: Consiste en hacer que el sistema aprenda a diferenciar entre la información relevante de la que lo no es. Y para ello utiliza técnicas de aprendizaje automatico y categorización de textos, y a diferenciar entre unas tematicas y sectores de otros.
* Rastreo automático: El sistema tecnológico rastrea periódicamente una selección de dominios y fuentes de noticias especializadas. Cuando se detectan nuevos contenidos, el sistema los descarga, categoriza automáticamente y selecciona los resultados relevantes, descartando los no relevantes. Estos contenidos actualizados y relevantes son las denominadas Alertas de Vigilancia. Éstas son almacenadas en el escritorio de la plataforma virtual y accesible para todos los usuarios, previamente registrados. Cada usuario recibirá entonces un conjunto de alertas específicas y personalizables según sus preferencias seleccionadas.
* Ranking y envío por correo: Según el medio de consulta y la frecuencia escogida, el usuario recibirá un resumen personalizado con las diez alertas más relevantes del conjunto de todas las alertas que el sistema tecnológico ha localizado, clasificado y priorizado. Para ello, se elabora un ranking automático por usuario sector de actividad, países de interés y temáticas escogidas en su perfil.

### Sistemas de análisis de sentimientos

## Capítulo 3.

## Integración de los sistemas de alerta y sistemas de análisis de sentimientos propuestos.

En el capítulo anterior se ha introducido conceptos como Internet, Web 2.0, Redes colaborativas y Redes Sociales, ahora se procederá a la extracción de la información utilizando el lenguaje R

### Primer apartado de este capítulo

### Segundo apartado de este capítulo

### Tercer apartado de este capítulo

## Capítulo 4. Recomendaciones o conclusiones.

Los capítulos intermedios servirán para cubrir los siguientes aspectos: antecedentes, problemática o estado del arte, objetivos, fases y desarrollo del proyecto.

En el capítulo 1 se describió bala, bla, bla.....

## Conclusiones y líneas futuras

Este capítulo es obligatorio. Toda memoria de Trabajo de Fin de Grado debe incluir unas conclusiones y unas líneas de trabajo futuro.

## Summary and Conclusions

This chapter is compulsory. The memory should include an extended summary and conclusions in English.

### First Section

## Presupuesto

Este capítulo es obligatorio. Toda memoria de Trabajo de Fin de Grado debe incluir un presupuesto.

### Sección Uno

|  |  |
| --- | --- |
| Tipos | Descripción |
| AAA | BBB |
| CCC | DDD |
| EEE | FFF |
| GGG | HHH |

|  |  |
| --- | --- |
| Tipos | Descripción |
| AAA | BBB |
| CCC | DDD |
| EEE | FFF |
| GGG | HHH |

Tabla 7.1. Tabla resumen de los Tipos.

## Apéndice A. Título del Apéndice 1

### A.1. Algoritmo XXX

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*

\* Fichero .h

\*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*

\* AUTORES

\*

\*

\* FECHA

\*

\*

\* DESCRIPCION

\*

\*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

### A.2. Algoritmo YYY

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*

\* Fichero .h

\*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*

\* AUTORES

\*

\*

\* FECHA

\*

\*

\* DESCRIPCION

\*

\*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

## Apéndice B. Título del Apéndice 2

### B.1. Otro apéndice: Sección 1

Texto

### B.2. Otro apéndice: Sección 2

Texto

## Bibliografía

1. "Internet." Britannica Academic, Encyclopædia Britannica, 15 Mar. 2016.
2. Flores Cueto, J. J., Morán Corzo, J. J., & Rodríguez Vila, J. J. (2009). Las redes sociales. *Universidad de San Martín de Porres*, 1-15.
3. Prato, Laura Beatriz: *“ Aplicaciones Web 2.0: redes sociales ”* [987-1727-13-5; 1-4492-7744-6] p.1. (2010).
4. Cabero Almenara, J., Barroso Osuna, J. M., Llorente Cejudo, M. D. C., & Yanes Cabrera, C. M. (2016). Redes sociales y Tecnologías de la Información y la Comunicación en Educación: aprendizaje colaborativo, diferencias de género, edad y preferencias. *RED. Revista de Educación a Distancia*, *15*(51), Artic-1.
5. COBO, C. & PARDO, H. Planeta Web 2.0. Inteligencia colectiva o medios fast food (Grup de Recerca d'Interaccions Digitals, Universitat de Vic. Flacso México.). Barcelona/México DF. (2007).
6. Prato, Laura Beatriz: *“ Aplicaciones Web 2.0: redes sociales ”* [987-1727-13-5; 1-4492-7744-6] p.16. (2010).
7. Freire, E. S., Ordóñez, F. S., & Martínez, M. P: “Redes sociales, un mecanismo de impulso del emprendimiento turístico en la provincia Tungurahua, Ecuador/Social networks to promote the tourism project in The Tungurahua Province, Ecuador”. *Retos Turísticos*, *15*(3). (2017).
8. Ponce, I. "MONOGRÁFICO: Redes Sociales-Clasificación de redes sociales." (2016).
9. Ponce, I. "MONOGRÁFICO: Redes Sociales-Clasificación de redes sociales." (2016).
10. Silva, Jorge Manuel das Neves, Soares, António Manuel Lucas, & Universidade do Porto. Faculdade de Engenharia. Gestão de informação de desempenho em redes colaborativas de organizações. (2012).
11. Franco Pereyra, R., Ortiz Bas, Ángel, Lario Esteban, Francisco Cruz, & Universitat Politècnica De València. Departamento De Organización De Empresas - Departament D'Organització D'Empreses. *Arquitectura, Metodología Y Plataforma Tecnológica Para La Ingeniería Y Operación De Redes Colaborativas. Una Aproximación Basada En Servicios Digitales*. P 54-56. (2012).
12. ]BROOKSHEAR J. Glean. Teoría de la computación Addisson Wesley iberoamericana Wilmington Delaware 1993
13. Gelbukh, Alexander. "Procesamiento de lenguaje natural y sus aplicaciones." *Korpus Sapiens. Sociedad Mexicana de inteligencia artificial* 1 (2010).
14. Cortez Vásquez, A., Vega huerta, H., Pariona Quispe, J., & Huayna, A. Procesamiento de lenguaje natural. Revista De InvestigacióN De Sistemas E InformáTica, 6(2), 45 - 54. (2014).
15. https://www.rstudio.com/products/rstudio/
16. <https://www.r-project.org/about.html>
17. Gonzalez, Andres, Gonzalez, Silvia. “An Introduction to T”(2000).
18. <http://www.ovtt.org/informacion_alertas>